Impact of Warping vs Smoothing for Time Series Similarity

Frank Höppner

Ostfalia University of Applied Sciences
Dept. of Computer Science, D-38302 Wolfenbüttel, Germany

Introduction. When dealing with time series, the application of a smoothing filter (to get rid of random fluctuations and better recognise the relevant structure) is usually one of the first steps. In the literature on time series similarity measures, however, the impact of smoothing is not explicitly or systematically considered – despite extensive experiments in, e.g., [2]. Instead, complex similarity measures are frequently applied (e.g. dynamic time warping (DTW)), which implicitly deal with noise, but mainly with temporal dilation and translation effects. So up to now it is unclear, to what extent the good performance of DTW is due to its smoothing or warping capabilities.

Optimal Filter. In this work we consider a simple Euclidean distance applied to preprocessed (smoothed) time series. It is unlikely that one similarity measure fits all problem types (or data sets), so by choosing an appropriate filter, we may adopt to the problem at hand. The filter is automatically determined given a training set of classified series, such that distances between series of the same (different) class are minimised (maximised). The obtained similarity measure is then tested in cross-validated 1-NN classification for various data sets (as in [2]) and compared against the DTW performance. Starting from Euclidean distance (without any preprocessing) as a baseline, it turns out that for many data sets a substantial fraction of the performance improvement obtained with DTW is also obtained by choosing the appropriate filter. In some cases, the performance is even better than with DTW, which is due to the fact that a filter is a versatile tool: for some problems it may be advantageous to distinguish time series by their derivative rather than the original series and in such cases a filter that estimates the derivative may be retrieved. For further details the reader is referred to [1].

References

1. F. Höppner. Optimal filtering for time series classification. In *Proc. 16th Int. Conf. Intelligent Data Engineering and Automated Learning*, 2015.
2. X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2):275–309, Feb. 2012.